Making room for emergence

Andrea Zhok

Abstract: We try to provide in outline an understanding of emergent properties, which should possibly make the idea of emergence not just plausible but compelling. It is our conviction that the core truth of emergentism is neither especially exotic nor counterintuitive, while its apparent eccentricity is essentially due to some prejudicial ontological assumptions. In the first half of the paper our argument develops through Jaegwon Kim's rejection of emergentism. We argue that Kim's use of both the "causal inheritance principle" and the "causal closure principle" in his criticism of emergence is unwarranted. In the second half of the paper we develop a positive account of emergence through a restoration of the ontological notion of quality. We contend that any monistic ontology, in order to account for experience, must make room for irreducible qualities and that efficaciousness cannot be denied to them. The novelty of emergent properties amounts to a priori unpredictability, due to the very nature of combination. Their efficaciousness is interpreted in terms of qualifying thresholds modulating the mode of efficaciousness.

Keywords: quality; efficient causation; formal causation; supervenience.

1. *Introduction*

The theory of emergent properties, which enjoyed its heyday at the beginning of the Twentieth Century,[1] after being in disgrace for more than half a century (McLaughlin 1992; Sperry 1991), has become object of renewed philosophical interest in the last decades (Beckermann *et al.* 1992; Clayton & Davies 2006). Emergentist ideas are strategically positioned in the context of contemporary reflection, since they try to reconcile the widespread acceptance of monistic materialism with the claim of the irreducibility of life and consciousness to classical physicalistic paradigms. Yet, although the reasons of interest for

[1] The theory of emergent properties has been mainly developed in the framework of so-called British Emergentism, whose major contributions have been Henry Lewes' *Problems of Life and Mind* (1875-9), Samuel Alexander's *Space, Time and Deity* (1920), Conwy Lloyd Morgan' *Emergent Evolution* (1923) and C.D. Broad's *The Mind and Its Place in Nature* (1925).

philinq VII, 2-2019, pp. 9-33
ISSN (print) 2281-8618-ETS

Submitted: March 2018
Accepted: November 2018

emergentism are easy to see, emergentist theses have been often regarded as tainted by irrationalistic implications.

In the following we try to provide the outline of a conceptualisation of emergent properties, which should make the acceptance of the idea of emergence unproblematic and its main features intuitive. In the first half of the paper (§§ 2-4) our argument develops through a criticism of Jaegwon Kim's rejection of emergentist ideas. It is our conviction that the core truth of emergentism, far from being anything esoteric, is easily available, once some misleading ontological assumptions are removed, and the refutation of Kim's account allows to expose those assumptions. In the second half of the paper (§§ 5-7) we will try to develop a positive account of emergence through a restoration of the ontological notion of quality and against the background provided by the problems of scientific predictability and physical causation.

2. *Outline of the idea of emergence*

Neither British Emergentism nor the most recent emergentist developments belong to a "philosophical school", and therefore they do not share any clearcut theoretical orthodoxy. Nevertheless it is possible to identify a set of ideas lying at the core of most accounts of emergence. We can isolate the following four characterizing traits:

1) *Naturalistic Monism.* Emergentist theses share with modern science a general ontological background, which is materialistic, but not physicalistic. As Clayton notices (2006: 2), the mention of physicalism turns out to be inappropriate, since there is no need in an emergentist framework to endorse just the specific conceptions of matter, where *physical* categorization is applicable. Assuming the epistemic priority of physical categories may pave the way to reductionist accounts.

2) *Supervenience.* Emergentism is usually concerned with properties,[2] with particular reference to properties that emerge from the "fusion", "synthesis" or "configuration" of parts into a whole. As O'Connor points out (1994: 97-98), the relation between the properties of the parts and the emergent properties of the whole should be understood in terms of *supervenience.* A property, or set of properties, *supervenes on* other (subvenient) properties if it is, to some extent, *different* from the subvenient properties, while being *determined* by them. If the subvenient properties are fixed, then the supervenient properties are univocally determined, but the reverse is not true: the same supervenient

[2] In the following we will understand "property" as characterised by *efficaciousness* of a kind (properties are as properties do) and by instantiability (properties can be instantiated by individuals).

property can be produced by a plurality of subvenient properties. Thus, supervenience is consistent with a functionalist understanding of the relation part-whole in terms of *multiple realizability*. The implicit mention of higher and lower levels, suggested by the use of the prefixes "super" and "sub", must be taken in a loose way. Subvenient properties can be regarded as lower in the sense of epistemologically "more basic", as well as in the sense of "belonging to a smaller dimension" (micro-macro relation; see Kim 1998: 92).

3) *Irreducibility.* Emergent properties should be irreducible to the properties from which they emerge. However, the mention of irreducibility does not contrast with any specific reductionist account: the claim of irreducibility in emergentism antedates all classical accounts of reductionism (Carnap, Morris, Neurath 1938-1969; Nagel 1961); "irreducibility" was expressed as non-deducibility (Broad 1925: 59) or unpredictability of emergent properties. Lloyd-Morgan (1923) introduced the distinction between emergent and *resultant* properties, such that properties *B* are resultant from properties *A* if *B*'s occurrence can be foreseen (when A are instantiated) from the sole knowledge of *A*. On the contrary, if our knowledge of the properties *A* does not allow us to foresee the existence of properties *B*, the latter are called emergent. Yet, this crucial definition needs clarification from two points of view. First, when we talk of "*pre*dicting", it seems that we have in sight a temporal order of succession. However, it is not necessary to interpret the relation between *A* and *B* with reference to an unfolding process: the emergent properties can be *co-existent* with the properties whose knowledge makes the relevant "prediction" possible. This leads to the second point, which concerns the relation between epistemic and ontological interpretations of emergence. Talking of irreducibility in terms of unpredictability or undeducibility is ambiguous as to its epistemic or ontological meaning. On the one hand, emergent properties claim to be ontological entities, on the other, by defining emergence with reference to "unpredictability", or kin notions, it seems that we are invoking an epistemic criterion (the subjective inability to predict). Since the kind of emergence that is mostly of interest is the ontological variety, which is not meant to depend on contingent subjective knowledge,[3] the unpredictability of emergent properties is often defined as "*unpredictability in principle*".

4) *Downward Causation*. The last essential feature required by emergent properties is that they must be able to produce causal effects on the very ontological level from which they emerge. This is tantamount to saying that

[3] It must be noted that not all epistemic unpredictability depends on subjective contingent inabilities; as Bedau (1997; 2008) notices, there are instances of epistemic emergence which are not contingent, since they depend on explanatory incompressibility.

emergent properties must not be epiphenomena. The intuitive idea of epiphenomenon can be conveyed by Huxley's famous metaphor (see Huxley 1874), according to which epiphenomena are related to reality as the steam whistle is related to the locomotive: it occurs whenever the locomotive is propelled, but it does not contribute to the relevant propulsion. When it comes to the mind-brain issue, the question of epiphenomenality becomes the following: if the mind counts as an emergent property (or set of properties), then it must not unilaterally depend on the cerebral substrate (upward causation), but must be able to produce in its turn effects on the brain (downward causation).

The notion of emergent property that we want to argue for, involves all four conditions, and in particular it involves both irreducibility in the sense of *undeducibility* and "downward causation", such that *efficacious properties of the whole may be absent in the parts*. On the opposite side, Kim's rejection of emergence denies that irreducibility and downward causation are compatible. That is, he grants the existence of *novel irreducible* properties supervening on a substrate of physical matter, insofar as they are *not* causally efficacious; or, alternatively, he grants the existence of causally efficacious properties supervening on a physical substrate, provided that they are reducible. Kim reaches this conclusion by developing an articulate reasoning, which we are going to reconstruct.

3. *Jaegwon Kim's criticism of emergence*

Kim's rejection of emergent properties starts from a confrontation with Davidson's anomalous monism, which issues into a refutation of mental causation. Kim shares Davidson's monistic assumptions, such that all events are physical, while mental features supervene on physical properties. His argument against mental causation runs as follows: let us assume that a mental event M causes a physical event P. If we accept materialistic monism, we have to grant that physical event P must have a *physical* cause, otherwise we should assume that P is causally produced by non-physical events, which would entail a dualistic picture. This assumption is named "*principle of causal closure*" and states that "[i]f a physical event has a cause at *t*, then it has a physical cause at *t*"(Kim 2005: 15). Now, if we grant that event P is caused by physical event P*, and if we remember that, by hypothesis, P is caused by mental event M, it seems that we have to do with a case of overdetermination, where there are two sufficient causes for the same event. This can be at most an accidental condition, not a systematic one. But if we remember that mental features supervene on physical features, a different option comes to the fore: we need not consider separately a mental and a physical sufficient cause, since they are ultimately one and the same, a physical cause. This is called the "*causal exclusion argument*".

Kim does not deny that mental properties, as well as many other properties that we ordinarily experience, are *novel* properties that supervene on their physical basis. He easily admits that new features emerge from known ones. His point, however, is that in a materialistic framework we are entitled to attribute causal efficaciousness only to physical events. This point seems to be directly derivable from the principle of the causal closure of the physical world, but Kim argues for it by separately elaborating the idea of "functional reduction", different from the classical Nagelian model of reduction.

According to Kim, the classical model of reduction suffers from the fact that bridge laws, i.e. the laws that should enable to reduce the vocabulary of a theory to the vocabulary of a second more comprehensive theory, are merely inductive (contingent) connections (Kim 2006: 194-195; Kim 1992: 125-126). In this sense Nagelian reductionism does not really yield any "*conceptual reduction*", where the truthful conceptual content of the reduced theory is comprehended in the reducing theory (see Kim 1999: 7-8). This problem is supposed to be overcome by Kim's functional reduction, whose steps are illustrated as follows:

> Let *B* be the domain of properties (also phenomena, facts, etc., if you wish) serving as the reduction base – for us, these contain the basal conditions for our emergent properties. The reduction of property *E* to *B* involves three steps:
>
> Step 1: *E* must be *functionalized* – that is, *E* must be construed, or reconstrued, as a property defined by its causal/nomic relations to other properties, specifically properties in the reduction base *B*. (…)
>
> Step 2: Find realizers of *E* in *B*. If the reduction, or reductive explanation, of a particular instance of *E* in a given system is wanted, find the particular realizing property *P* in virtue of which *E* is instantiated on this occasion in this system;
>
> Step 3: Find a theory (at the level of *B*) that explains how realizers of *E* perform the causal task that is constitutive of *E* (i.e., the causal role specified in Step 1). Such a theory may also explain other significant causal/nomic relations in which *E* plays a role. (Kim 1999: 10-11)

In order to grasp the scope of the proposal of functional reduction, we must recall that classical reductionism entered in crisis, among other things, under the weight of the functionalist contention that the same property can be implemented by a multiplicity of different substrates. Kim's account of functional reduction apparently grants that properties can be multiply realised, and requires, in order for reduction to take place, just that the property to be reduced be defined in causal terms efficacious at the level of base properties. The level of base properties is for Kim the physical level. Once the property has been functionalised, we should search for appropriate realisers at the physical level;

finally, the connection between the causal functionalisation and the discovered realisers should be provided by an appropriate theory that "*explains how*" the realisers perform the causal task defined by the proposed functionalisation.

One of Kim's favourite examples of functional reduction concerns the reduction of genes to DNA. First, "genes" are defined in terms of their causal power to reproduce phenotypic traits across generations. Secondly, the causal efficacy of DNA in carrying those traits from one generation to the following is discovered in the realm of physical properties. Finally, "we have a theory that explains just how the DNA molecule is able to perform this causal work." (Kim 1999: 10) This procedure is supposed to reduce the apparent heterogeneity between second-order properties (like "gene") and first-order properties (the physical instantiations of deoxyribonucleic acid). Such heterogeneity has not been really reconciled by classical reductionism, because of the merely inductive character of bridge-laws, while functional reduction is supposed to provide a *theoretical* bridge that *explains how* the macroscopic property of reproducing phenotypic traits is nothing but a physicochemical property of a nucleic acid.

Still, bridging the gap between a general property and a particular physical substrate does not seem easy. Functional reduction states that if we recognise that property P is essentially characterised by causal power C and if we discover a physical substrate S to which C can be attributed, then we can posit P = S. But, as Ned Block (2011) notices, this passage is incorrect: let us assume that P is "being somniferous", which translates into: P is endowed with the causal power C of producing sleep. Then, if we find a physical substrate S capable of causing sleep (e.g., Phenobarbital), we should be entitled to say that P = S, that is, being somniferous equates with Phenobarbital. But this cannot be true, since there are many other physical substrates endowed with the same causal power (e.g. Diazepam). This difficulty cannot be overcome by claiming that, in fact, being somniferous should be equated not with a single substance, but with the disjunction of all sleep-inducing substances (either Phenobarbital or Diazepam or Lorazepam, etc.). This equivalence is mistaken because the *property* "being somniferous" refers to *infinite possible* embodiments of the sleep-inducing power, and never to *finite actual* embodiments. In fact, the general nature of a property, causal or not, is never equivalent to a list, not even an exhaustive list, of particular embodiments. It seems that the only way in which the ontological gap between the dimension of general properties and the one of particular embodiments can be reduced is by denying true ontological value to one of them;[4]

[4] It seems that, if reduction there must be, you must either deny the ontological value of general properties (as Kim does) or the ontological value of individual embodiments (which amounts to an idealistic solution).

and indeed, this is what Kim in the end proposes, by suggesting the ontological vacuity of functional properties.

This point is argued by means of what he calls the "*causal inheritance principle*", according to which "[i]f a functional property *E* is instantiated on a given occasion in virtue of one of its realizers, *Q*, being instantiated, then the causal powers of this instance of *E* are identical with the causal powers of this instance of *Q*" (Kim 1999: 16). In this sense, functional properties and in general second-order properties (among which mental properties) are considered void of any causal role different from the causal power of their realisers (first-order properties, i.e. physical properties). The core intuition that supports Kim's argument is that causality is a relation concerning particulars only, where particulars are physical. This picture is far from exotic: causality is supposed to be a relation between existing things and only individuals exist. Under these premises universals, ideas, general properties (and functional properties among them) are not ontologically efficacious, or rather, they have no efficaciousness distinct from the efficaciousness of their physical embodiments. As Kim states, the only justifiable role left for functional properties is epistemic, since they "may play a practically indispensable role in our discourse" (17-18), but they have nothing to contribute at the causal level.

Kim's criticism of emergent properties rests on the causal inheritance principle and the principle of the causal closure of the physical. In fact, on an appropriate understanding of the "physical", each principle alone is already sufficient to delegitimize emergence. The causal inheritance principle essentially states that all ontological efficaciousness[5] inheres in *physical realisers* and that such efficaciousness is just "inherited" by higher-order properties. This straightforwardly implies that, if we grant that physical properties are not emergent themselves, then there cannot be such a thing as a truly efficacious emergent property. The causal closure principle essentially states that everything that might ever have causal efficaciousness on physical events must belong to the realm of physical causes. This implies again that, if we grant that the realm of physical causes forbids novel forms of causation, then there is no ontological room for emergence. In both cases, the soundness of the argument depends on the meaning to be attributed to the *physical* realm and to the relevant efficaciousness. But what is meant here by "physical"?

As we said, Kim does not deny that macroproperties are causally efficacious, or that there are effects of macroproperties, which are not to be found at the

[5] We will use all through the text the expression "ontological efficaciousness" to express a most general notion of causality, which is not committed yet to any specific interpretation concerning *how* efficacious relations in nature must unfold (e.g. over time or instantaneously; by contiguity or at a distance; with perfect uniformity or not, etc.).

microlevel. According to him, macroproperties can be regarded as causally efficacious because they are *themselves physical or reducible to physical*. This can be maintained only by introducing an extremely comprehensive acceptation of the physical. Indeed Kim firstly defines all entities and properties of *basic physics* as “physical entities”, and then includes in the physical realm also any entity aggregated out of or composed by physical entities, and any second-order property defined over physical entities (Kim 1997: 294). This means that the realm of the physical goes well beyond the entities with which physics usually deals, by extending over chemical and biological entities, and over all entities identified by properties that are functionally reducible (296). This makes of the thus-defined physical dimension an up-to-date version of *res extensa*. That is confirmed by Kim’s reading of the principle of the causal closure of the *physical* world, which amounts to a claim of *materialistic* monism, excluding only supernatural (spiritual) causes. In the end Kim’s physicalism does not seem to differ from mainstream scientific naturalism, with just a moderate inclination to grant a greater explanatory power to physical categories. If this is the case, then the physical realm is just the realm of all objects and properties with which the sciences of nature are concerned.

But, if we accept such a permissive reading of the physical realm, then we apparently have a problem, because it becomes difficult to recognise any *specific kind* of *causal* relation, which should be able to account for all ontological efficaciousness within physics and between physical elements and chemical, biological or neurological units. Indeed, if we accept that there are physical causes that bring to light biological organisms, what conclusions should we draw about the *nature* of such causes? Should we say that physical causes “remain exactly the same” when they operate within the biological organism and before issuing into that organism? Or should we say that they have become a different sort of causal relations and that now we have to do with specifically biological effects? And how are we going to decide between the two options? The problem is that, if we do not have any clear-cut idea about the *nature* of ontological efficaciousness (“physical causality in itself”, in Kim’s reading), then we are *not* in a position to claim that instantiations of *prima facie* non-physical causation inherit “nothing more than” the causal powers of the physical basis. In fact, physics by itself is not especially committed to any particular understanding of causality, and actually physical laws usually need not refer to causal relations at all. And we certainly do not gain any stronger commitment to a specific reading of ontological efficaciousness, if we water down our understanding of the physical, by covering all natural sciences.

But before developing this point let us ask whether we can gather from Kim’s analyses, what his understanding of the workings of physical causality is.

4. *Is ontological efficaciousness intrinsically temporal?*

Kim makes explicit his understanding of ontological efficaciousness, when he denies the possibility that irreducible properties of the whole affect the properties of the parts of the same whole. What emergentist theses need, he says, is not just downward causation, but *reflexive* downward causation, that is, a kind of causation where wholes are able to modify the *way* in which their own parts operate. Kim's refutation of reflexive downward causation proceeds by separating two cases, respectively concerned with *synchronic* and *diachronic* effects. Here we are discussing only the first case, leaving the second one to the next paragraph.

In synchronic reflexive downward causation the possession of property M by the whole W at time *t* causes the possession of property P by parts of W at the *same* time *t* (Kim 1999: 28). The objection that is raised against this case is the following: property M can emerge in W, because at time *t* the parts of W are endowed with certain properties and have a certain micro-configuration. If this is the case, it may seem awkward to say that, at the very same time *t,* such properties of the parts are in their turn caused by the property M of the whole W. It seems that an object can exercise the causal power pertinent to its own properties only if it *already* possesses those properties. Therefore, an instance of reflexive downward causation seems to defy our imagination, by representing a system where the properties of the whole can subsist only if they are supported by the properties of the parts, which are in their turn simultaneously caused by the properties of the whole.

This objection is undefeatable insofar as ontological efficaciousness is interpreted in terms of efficient causality, with particular reference to the idea that causes must be antecedent to their effects. Under these premises, to conceive a whole that modifies the nature of the very parts that *simultaneously* determine its own existence is puzzling indeed. However, it must be noticed that it is unusual to conceive of the relation part-whole in strictly causal terms. There is certainly something odd in saying that a whole *causally* determines its own parts by means of a *non-temporal* relation, yet, one may wonder whether the oddness rests rather on the use of the ordinary notion of cause than in some alleged difficulty in conceiving the determination of parts by the relevant whole. As Craver and Bechtel notice (2007: 554), our ordinary notion of causality assumes the separation of cause and effect, because this is an operative condition necessary to carry out our experiments: we need first to introduce a modification and then to detect possible repercussions. But if we drop this requirement, we do not have any special difficulty in observing instances of apparent "symmetrical efficaciousness", where changes in the whole and changes in the parts *co-occur* in a systemic (organic) way. The point is not empirical but conceptual. If we observe an instance of spontaneous change (e.g., the biological "working" of living cells), we are never in a position to affirm that only elementary parts are truly efficacious, while the relevant whole is not, at the very least because, logically, any change in the parts is also a change in the whole.

Our ordinary notion of causation is not really able to capture ontological efficaciousness as such, but focuses on that subset of ontological efficaciousness that we can *steer*. As von Wright observed (1971: 63), sheer recording of regularities could never rationally justify the articulation of phenomena into causes and effects. Such articulation becomes properly possible only when we ideally *intervene* on the system and observe the repercussions correlated with our intervention. This cannot mean, however, that we have discovered how nature in itself exercises its efficaciousness; rather, our interference allows us to introduce in the investigated system a pragmatic order, which teaches us how some changes can be prompted. This pragmatic order dictates the articulation of ontological efficaciousness into an *active* component (cause) and a *passive* component (effect). The active component *must* be antecedent to the effect, otherwise we could not separate an active and a passive factor, and therefore we could not causally govern the passive factor. If something in nature displays a change in its manifest properties when it occurs in a different whole, we are likely to describe this change just as a "*phenomenon*" and not as an *effect*, because here it is hard to make out an active cause and a passive effect. But if we refrain from speaking of causation in these contexts, this seems to signalise a problem of how we like to interpret causes rather than to show the ontological inefficaciousness of the relevant circumstances. When we assume that effects propagate by contiguity rather than at a distance, and over time rather than instantaneously, we introduce requirements that are pragmatically sensible: if my intervention on a supposedly isolated system produces a simultaneous effect on Alpha Centauri, that is nothing of my operative concern. This means that there is a natural bias in favour of the idea that ontological efficaciousness must unfold in terms of temporal propagation. Thus, there are good pragmatic reasons for exploring the efficaciousness of nature in non-holistic and temporally articulate terms, and our experiments rightly make any possible effort to bring to light distinct effects within ideally isolated systems. Yet, we must be wary not to draw hasty conclusions on the nature of ontological efficaciousness as such from this propensity of ours. As we will see below (§ 7), there is a rational way to understand instantiations of ontological efficaciousness which need not be figured out as "propagation over time".

5. *What is inherited in the causal inheritance principle?*

The second part of Kim's argument against emergent downward causation is developed along lines apparently similar to the above mentioned "causal exclusion argument". *Diachronic* downward causation should be a relation such that the possession of property M by the whole W at time *t* causes the possession of property P by parts of W at time $t^1(>t)$. Yet, emergent property M, which is supposed to produce property P in the parts of W at t^1, must supervene on basal conditions P*. But now we have apparently to do with two possible claims of causal sufficiency relative to P, which by definition should be caused by M, while P* is sufficient to determine M. If this is the case, according to Kim the only sensible solution is to admit that all ontological efficaciousness belongs to the basal (physical) conditions, so that P* must be considered the true cause of P and displaces M as a true cause. This does not deny the subsistence of instances of downward causation from M to P, but interprets all such instances as reducible to elementary (physical) causation (Kim 1999: 32).

Although the form of this argument reminds of the causal exclusion argument, in fact it rests on a slightly different basis. The core reasoning can be summarised as follows: since an alleged emergent property of the whole supervenes on a physical basis, it is reasonable to think that the supposed effects of the emergent property are actually effects of the subvenient basis. The causal exclusion argument relied on the causal closure principle to the effect that, if we assume the *explanatory exhaustiveness of physical causes*, then supervenient mental properties have nothing left to do. Instead, the argument against diachronic reflexive downward causation rests rather on the causal inheritance principle: the supervenient property of the whole must *inherit the causal powers of its subvenient basis*, as all second-order properties must inherit the causal powers of the relevant first-order properties. In this criticism, it becomes explicit that for Kim supervenient properties cannot belong to the sphere of real efficaciousness. They are not properly "caused" by subvenient properties since supervenience in itself is no causal relation, and they must not add causal features to the picture, because otherwise we would have "new ontological entries" in our physical universe, which would mean that we admit of "supernatural" powers, dropping naturalistic monism. As above, either a novel property is reducible to physical properties (and then its efficaciousness can be granted), or it is not reducible to physical properties (and then it must be an epiphenomenon, like *qualia*). Here we are approaching the heart of Kim's argument, where inference yields the ground to intuition. He claims that supervenience implies a *commitment to physicalism* (Kim 1999: 14-15). But even if we grant a permissive reading of the physical, and therefore of what counts as physicalism, this claim remains doubtful. If he wants to state that properties that do not look like physical properties (e.g., mental ones) are anyway anchored in the physical sphere, this is consistent with the definition of supervenience and can be granted. Yet, in such commitment to physicalism Kim includes also the idea that supervenient properties do not bring to light novel causal powers, but *inherit the causal powers of the subvenient properties* (which are physical properties). This is a much stronger thesis, which implies, among other things, that ordinary physics, or even the sciences of nature as such, are committed to a specific idea of what *physical causation* is. This is certainly Kim's view, when he contrasts his proposal of functional reductionism with classical reductionism. He maintains that functional reduction could perform actual theoretical reductions, which would not suffer from the problems characteristic of classical reduction: true theoretical reduction is supposed to provide not just inductive bridge laws, but *conceptual equivalencies* between the terms of the theory to be reduced and the terms of the reducing one. And a *conceptual* reduction must be something able to *explain how* physical realisers perform the causal task defined by the functionalisation of the relevant property. Yet, the idea of "explaining how" physical causality works is definitely opaque. Physics is not especially committed to any particular account of causation, and laws of nature are mostly symmetrical as to temporal order and do not make any reference to causes. This does not mean that physical *praxis* can do without the notion of cause, but physics can do very well just with a Humean account of causality, where causes do not have any special identity separable from the identity of the relevant antecedents and consequents: here no specific essence of causality comes to the fore. Indeed, whenever in Kim's texts we come close to the disclosure of what a specific *conceptual* reduction should amount to, we find curiously evasive expressions. For example, when Kim exemplifies functional reduction with reference to the reduction of genes to DNA, he states in the end that "we have a theory that explains, *at least in broad but persuasive outlines*, how the DNA molecule is able to perform this causal task" (Kim 1997b: 51, emphasis mine) or, in another occasion, that "[w]e *presumably have a story* at the microbiological level about how DNA molecules manage to code and transmit genetic information" (Kim 1999: 11, emphasis mine). The vagueness of these expressions are probably meant as suggestions that the relevant scientific knowledge is still very much in progress. However, the point is that even if we had a closer look at the relevant microbiological description we would never meet any spot where, so to say, the causal connection "displays its interiors" so that we could grasp the *ultimate reasons* governing the production of effects. Such a conjectural dimension where causality becomes conceptually transparent is purely mythological. Scientific knowledge of causal relations is *more detailed* than pre-scientific knowledge, and this is quite important, since it provides a plurality of "handles" where we can ideally intervene to manipulate causal outcomes. Yet, regardless of how detailed the analysis is, we are never in a position to discover more-than-inductive reasons *why* phenomena of a kind generate phenomena of another kind. The only ideal possibility to bridge such conceptual discontinuities would be to discover a micro-level where all heterogeneity between causes and effects dissolves, but, as we will see, this option is properly unconceivable.

This seems to be a blind spot in Kim's arguments: the sciences of nature are not committed to any specific idea of causality and are therefore perfectly compatible with emergent properties. If Kim wants to argue against emergence, he cannot rely on a settled uncontroversial model of ontological efficaciousness, but should explicitly argue for one. Indeed, only if we had such a specific model before our eyes, we could understand what the causal closure principle really *forbids*, by implying that physical events have *specifically physical causes*. And only in the presence of such a model we could see which *limits* are posited by stating that second-order properties *inherit nothing but* the causal powers of first-order (physical) properties.

Yet, this "blind spot" is not something peculiar to Kim's arguments and it goes to his credit that his argumentative clarity helps to bring those assumptions to light. In fact, doubts about the acceptability of emergent properties depend on a widespread tacit understanding of how physical causality (ontological efficaciousness) should work. Let us try to understand features, grounds and implications of that tacit assumption.

6. *Causality and the conservation of quantity*

To make explicit the intuition that guides Kim's implicit understanding of ontological efficaciousness, we should ask what is forbidden by the principles of causal closure and causal inheritance respectively. In fact, they seem to exclude pretty much the same prospect: they forbid that new ontological entities pop up in the natural system of causal relations. If we concede that entities endowed with causal properties may emerge unexpectedly, it may seem that unexpected effects could be ubiquitous, the conservation laws of physics would be threatened and materialistic monism should be abandoned.

Some of the most authoritative recent interpretations of causality in nature propose that causal interaction be understood as exchange of a conserved quantity (like energy or momentum; see Fair 1979; Salmon 1998; Dowe 2000). This model captures a significant aspect of the scientific treatment of causal relations, where an important role is played by the ability to follow the preservation of some physical quantities (especially energy) along a process. On the other hand, it should be also obvious that this idea does not capture our general intuition of causal efficaciousness, which does not depend on any quantitative evaluation and much less on the detection of a preserved quantity. If I shout to get attention, I am inclined to think that my voice was efficacious in producing your attention, quite regardless of any concern for the preservation of physical quantities. Still, the idea is that deep down, at the core of all truly efficacious causality there must be the transmission of a physical quantity. But why does the identification of a *quantity* look so prominent? The reason is apparently rooted in the classical scientific tradition and has to do with the forecasting aims, which nourish our very interest for causality.

Quantification lies at the roots of modern science and allows the exact registration, intersubjective control and technical re-instantiation of (apparent) causal nexuses. But here quantification is a methodological instance, which in itself cannot support any specific ontological vision. Yet, historically such a strict separation between methodological and ontological instances has not been much respected. The ideas of Greek atomism have been revived by the birth of modern physics, and the traits that Democritus attributed to atoms have been translated by Galileo into so-called primary qualities. Conceiving nature as essentially represented by primary qualities is what allows Galileo to state in the *Assayer* that the grand book of nature is written in the language of mathematics. The powerful core of this proto-physicalistic vision is the intrinsically *quantitative* character of nature, which provides a strong justification to our aspirations to essential knowledge and technical manipulation of the world. In the Galilean universe we are supposed to have to do with unchanging quantifiable elements mutually related by mathematizable relations. And the guarantee that elements are quantifiable and relations mathematizable rests on the assumption that "deep down" elements are intrinsically *quantitative* and relations are of a *mathematical* kind. However, it should be clear that the applicability of mathematics to the world does not require nature to be intrinsically akin to mathematics. The subsistence of regularities in the unfoldment of properties is quite enough to allow the fruitful application of mathematical devices, which are in themselves highly flexible and able to grasp a wide plurality of correlations. Units of measurement are introduced in virtue of their ability to provide constant results over repeated measuring acts and they make possible the *quantification* of worldly *qualities*, insofar as such a quantification is feasible. We do not need straight lines and triangles to exist in nature in order to measure length by rigid yardsticks and to calculate topographic distances through trigonometry.

Although this point should not be especially controversial, the tacit assumption that nature must have "deep down" quantitative character is quite resilient and widespread. Yet, in order for this idea to remain operant, it seems crucial for it not to be stated too openly, since it is far from easy to figure out what an "essentially quantitative" world should be like. Should we imagine a universe constituted by perfectly simple "number-like" atoms? Or a Democratean universe composed of atoms differentiated just by their geometrical shape and size? I do not know of many who would explicitly support similar views. However, if we do distance ourselves from these highly speculative theses, we must ask: what *else* could ever support the idea that the union of two properties *always* causally produce *homogenous* (*congenerous*) properties? But this is precisely the idea that in the natural world there is only room for "resultants" and never for "emergents". This picture is plausible just in a universe composed by elements which are connected only by quantitative relations, that is by relations that *preserve unmixed and unchanged the identity of the properties of the related elements*. If a plurality of elementary qualities is admitted, their interaction must not be qualitative but quantitative, that is, their properties must remain juxtaposed to each other and the range of their variation must concern just the degree to which identical properties add up producing congenerous effects. If, on the contrary, we accept that the relations between elements of our universe *need not* preserve the identity of the properties of the related elements, then we must grant that the interaction of elements can bring about novel properties, which is what defines emergence.

However, why should we reject the idea that the properties of our world can be accounted for just by relations of "additive" kind, operating over a set of elements with original and unchanging properties? Strictly speaking, there are no mandatory reasons to reject this picture, but we should be aware of its implications. If we assume a monistic ontology, to preserve that thesis we must grant that *there are at least as many basal ontological properties* as there are *experienced* properties. The reason is the following. When we contrast appearance and reality while experiencing properties, we explain unreal appearance by assuming that subjects somehow "project" delusional properties on reality. However, in a monistic ontology *even a delusional property* is a property that belongs to reality (maybe as an event in the brain) and in principle we must account for its existence. The possible delusional aspects concern the *actuality* of the *relations* that I attribute to some apparent properties, but not the existence of the phenomena: I can wrongly believe that the experienced flash causes the experienced thunder, but the distinct experiences of the flash, the thunder and the impression of a causal connection must all have a place in our ontology. However, if no property can be newly generated in my universe, then for each experienced property there must exist basal ontological properties supporting that experience. This option is not intrinsically absurd, but suffers from an extreme ontological profligacy.

In substance, if we want to account for the properties that we experience, we must provide a picture where all phenomenal properties have an ontological ground, and this implies that such a qualitative plurality either is assumed at the level of ontological elements, or is generated by the qualitative interaction of ontological elements. *In neither case we are in a position to evict quality from ontology*. Let us try and see what happens if we begin to take seriously the idea that quality is ontologically fundamental.

7. *Quality and predictability*

It is remarkable that, although the philosophical notion of quality is among the most ancient and revered, its philosophical treatment has been rather scant, sometimes controversial (Aristotle) and often disparaging. Aristotle's treatment of quality is notoriously obscure, but is also probably the only authoritative attempt to articulate the idea while preserving its specificity. The Galileian, Cartesian and Lockian treatments of qualities are mostly devoted to subjectivise them and to expel them from ontology, where only the quantifiable aspects of quality (primary qualities) should appear. It is noticeable that Kant's discussion on quality in the *Anticipations of Perception* of the *Critique of Pure Reason* regards quality just as the "reality of appearance" (sensuous δόξα), which can enter scientific reality only according to its *intensity*, which is quantifiable (Kant 1998: 290-295; KrV A166-176/B208-218). It is no less remarkable that Hegel in the *Science of Logic* (Hegel 1999: 157ff., vol. I, book 2, section 3) substantially accepts Kant's ontological subordination of quality to quantity by arguing that quality obtains proper rational position only in the form of *measure*. In all rationalistic treatments, quality is treated as mere hindrance for the intellect, as something to be quickly overcome and forgotten *precisely because of its resistance to analysis*. This attitude is understandable, but also dangerously limiting when applied to ontology, because it induces to neglect or underplay some universal features of the world we experience.

When we consider ontological properties in quantitative terms we treat them as if they were always intrinsically prone to be submitted to algebraic operations. Yet, at least the appearance of phenomena often suggests an unaccommodating "algebra of qualities". If in ordinary algebra $a + b$ makes sense only if both a and b are reducible to a congenerous basis, in a "qualitative algebra", consistent with many natural appearances, the union (fusion, configuration, etc.) of two qualities generates a third quality, which *may have* *different properties from the constituting qualities.*[6] Insofar as ontological elements are qualities, we cannot a priori adopt logical or conceptual rules to anticipate what the result of the union of two or more qualities is. Any composition of qualities prima facie resists being treated in additive or deductive terms: we cannot add or subtract different qualities, nor can we deduce the consequent properties. The main principle of an "algebra of qualities" should look like: $\alpha \,\&\, \beta = \gamma$, where *different* qualities are named by the Greek letters and their *non-additive union* is designated by &. This is the "unruly rule" according to which chlorine (a toxic gas) and sodium (a soft metal) end up yielding table salt (Rothschild 2006: 152-153), or according to which the Jamaican sweet orange and Indonesian pomelo yielded grapefruit. This "rule" cannot depart from recorded experience and is refractory to any unqualified application of mathematical devices.

Nevertheless, emergence does not exclude *a posteriori* predictability. Insofar as natural processes display uniformity, we can "tame" different qualities by the choice of appropriate yardsticks (apples and pears can be summed as fruits) and we can reliably infer properties in the wake of settled experiences (the union of hydrogen and oxygen under certain conditions and proportions ensures the production of a transparent, thirst-quenching liquid). Sagacious use of well-chosen *qualitative samples* may allow the quantification of phenomena, which consists in the institution of a *ratio between a reliable qualitative sample* (unit of measurement) and *objects of a class suitable to be thus measured*. Under this perspective quantification looks like a legitimate but secondary procedure, which presupposes the existence of original qualities and exploits the peculiar traits of some of them (rigidity, regularity, stability, etc.).

In such a framework *emergence* of properties represents just the natural background, from which we may occasionally learn to extract qualities that are available for quantification (primary qualities and the like). This kind of emergence does not have any of the connotations of creation or miracle, which have often discredited emergence, since *repeatability* of both preconditions and outcomes is mostly possible. This means that we are not concerned with what van Gulick calls "radical kind emergence" (Van Gulick 2001: 17), which would be a relation where the emergent property is not just different from the properties from which it emerges, but is also *not dependent on them in any necessary or law-like fashion*. This kind of emergence, while being marginally tolerable, cannot be anything but a rare option in our world, since it conflicts with the uniformity of natural processes. It must be emphasised, however, that this understanding of

[6] We are using here the term "quality" as shorthand for "entity or substance endowed with qualitative character (i.e. with 'irreducible property/ies')".

emergence is foreign to the usual idea of emergence, which is traditionally exemplified by regularly produced instances like the transparency of water. We have ordinarily to do with properties that are prima facie *efficacious*, and are *novel* in comparison with the properties of the qualities from which they emerge.

However, even if we adopt a qualitative ontology, could we not work out a view supporting not just *a posteriori* but also *a priori* predictability? Could we not suppose that the novelty of emergent properties is merely apparent, depending on our epistemic inability to deduce emergent properties from the separate knowledge of their basal qualities (see O'Connor 1994: 98)? Even if new properties are not tied to basal qualities by deductive (logical or mathematical) relations, we could speculate that, if we had God's knowledge of all basal qualities, each state of the universe could appear strictly derivable from the immediately previous stage. Even if for a human intellect complex properties can be *anticipated* only if (and insofar as) qualities are quantified, one may conjecture that superhuman capabilities having perfect knowledge of all *elementary qualities* could nevertheless "simulate" in advance all steps leading to the knowledge of all possible properties and states of the universe, even if they are actually novel. Here the *novelty* of emergent properties would not be delusional, but could be nevertheless anticipated by a mind capable of *individually* inferring each step from the previously occurring qualities. This view makes us contemplate the possibility that the novelty produced by the union of qualities could be anyway interpreted, in principle, as a resultant.

But not everything is clear in this hypothesis. In particular, in which sense are we justified in speaking of "simulating" or "inferring" the new qualitative states of affairs? If we grant that actually *novel* properties turn up, can we still regard the anticipatory process leading to them as a *cognitive* process (as simulations and inferences are supposed to be)? Actually, it seems that in order to learn the novel qualitative outcomes of such process, the divine intellect should *perform* it, since in the absence of the recursion of identical properties this process cannot be cognitively compressed (see Bedeau 2008). At this point, it seems that we can formulate only two (highly speculative) options.

First, we could suppose that in order to "calculate" the emergence of properties the Divine Intellect should just run the world as it is. But this would no longer be any anticipation and even God should just wait and see what the future holds.

Second, we could imagine that such Divine Intellect, like a superhuman experimental scientist, could perform ontological simulations in vitro, that is, could test an exhaustive set of basal qualities by separately running their combination. In order to support a priori predictability, this hypothesis should assume that from an exhaustive combination performed over a selection of basal properties (like Mendeleev's elements) all properties (even emergent ones) could be discovered. This hypothesis, however, can be viable only by assuming that recursive occurrences of the same property in the same process cannot produce new properties over the process. But from the above mentioned principle of our "algebra of qualities" we should immediately draw an incompatible corollary. Indeed, if α & $\beta = \gamma$ (as defined above), then it is to be expected that recursively γ & $\beta = \delta$; δ & $\beta = \varepsilon$, etc., where γ, δ, and ε are further novel properties. That is, the recursive occurrence of the same property (β) in a different stage of the same process can generate further novel properties.[7] If we take seriously the idea that the union of two qualities can generate a novel quality, then the sheer number of new instantiations of the union of a finite number of qualities could always generate newer and newer properties. Therefore the very fact that, by hypothesis, the experimental set on which our "superhuman scientist" operates is quantitatively circumscribed would deny full cognitive access to some qualitatively emergent properties.

8. *Qualitative efficaciousness and ontological thresholds*

The last paragraph should have shown that, in an ontology governed by qualitative relations, *a posteriori predictions* are possible, but *a priori* predictions are not. There is however a last piece of argument missing, if we want to make truly room for emergence. As we have seen, Kim's rejection of emergentism targeted properties, which should be at the same time *irreducible* to physical properties and *truly efficacious*. Irreducible novelty was granted to properties, if causal powers were denied to them: this is the field of epiphenomena. Kim is not certain about the extension to be attributed to such epiphenomenal area: mental *qualia* are certainly included, mental properties are included to the extent that they are not conceptually reducible to physical accounts. But if the condition for not being epiphenomena is the reducibility to physical accounts, prima facie we have to include in the area of epiphenomena *all* mental properties, insofar as no mental property has ever been *conceptually reduced* to physical accounts as yet. The prospect that the advancement of science will in the end produce such accounts is an interesting hypothesis to entertain, but is no argument for excluding mental properties from the field of epiphenomena, according to Kim's requirements.

However, as we said above, alleged epiphenomena cannot be foreign to ontological efficaciousness. On the one hand, they must be *produced* (caused to

[7] This conjecture becomes less abstract if we consider that natural sciences like biology bear witness to the fact that the same factors (e.g., enzymes) can generate quite different effects when they occur in different contexts or at different stages of the same process (Soto, Sonnenschein 2006: 64).

be); on the other, they must be at least *efficacious enough to mislead* us. The option that remains open is that *qualia* and mental properties in general, while efficacious, do not have the kind of efficaciousness that they pretend to have (i.e., their content is generally false). Yet, this reading of epiphenomena is untenable. If mental properties are epiphenomena, because only physical causes are truly efficacious, then no mental relation is what it claims to be: volitions do not produce actions, inferences do not logically guide consequences, percepts are not drawn from reality, motivations do not drive our will, etc. Our most fundamental methodological and epistemological instances, like our predilection for simpler and more comprehensive rational solutions would be just delusions motivating further delusions. Under these premises, as Husserl famously argued (1913: §§ 32-38), the outcome is a complete breakdown of any claim to truthfulness, inclusive of the very contention that the only true dimension of efficaciousness is the physical one.

Still, even if we are finally convinced that our ontology must contemplate irreducible qualities, and that the apparent efficaciousness of those qualities must be mostly *actual* efficaciousness, the main intuition underlying Kim's arguments remains standing. Kim's exclusion argument and his criticism of downward causation rely on the core idea that if physical (natural) effects have physical (natural) causes, then we cannot invoke further non-physical causes without risking to destroy the whole project of natural sciences, because non-physical causes could conflict with and exauthorate physical causes. But this threat would be actual only insofar as causation intrinsically requires the transferral and preservation of physical quantities: indeed, if all causation is just the transferral of a physical quantity, then any addition of causes to an exhaustively described physical system would be unintelligible. In order to confront this idea we have to provide an alternative view of ontological efficaciousness.

Now, if we rely on what is offered by experience (scientific or not, truthful or illusory), we should easily recognise the ubiquitous presence in the observable world of *selective thresholds* or *formal discontinuities*. By that we mean the constant appearance that not everything that happens or changes at a certain level, or in a certain entity, produces events or changes at a "higher" level, or outside the relevant entity. We recognise, for instance, that not everything that *happens in* an atom, or a cell, or a living organism, or a planet, etc. produces effects outside the atom, the molecule, the cell and the planet respectively. It is precisely such kind of discontinuity that allows speaking of the threshold between "inside" and "outside" of the relevant items. We make experience of organisms, which change physical constituents over time without losing identity; we detect transitions of energy levels in atoms that remain the same atoms; we identify molecular oscillations, without molecules turning into different molecules; and arrows can be thrown against the sky, without anything in the superlunar sphere being concerned by that. In a world devoid of such discontinuities each (putative) elementary event would fluidly and continuously spread its effects without boundaries and no entity endowed with identity could *appear*. It must be stressed that under the present premises, it is wholly irrelevant whether such discontinuities are judged to be "existent in themselves" or just to be an "epistemic effect" (due to the way in which subjects organise experience, or the like). Once more, if such discontinuities are available to consciousness, in a monistic framework they are part of ontology, even if their specific place in our ontological vision may be uncertain. Here we are not interested in ascertaining that discontinuities reside in "things in themselves" or in the relation between things and ourselves; in any case the subsistence of such essential discontinuities cannot be denied. Such thresholds, if you like, could be considered the intuitive correlate of what physics has successfully explored under the category of *quanta*. Yet, any physical technicality is here beyond our concern: the point is just that we have binding reasons, immanent to consciousness, to assert that ontological efficaciousness (naturalistically: matter/energy) expresses itself by essential discontinuities.

That said, we have to observe that the existence of such discontinuities implies possible *asymmetries of causal transmission*, whenever a propagating effect comes to a *qualified* threshold. Small changes reaching relevant thresholds can have macroscopic effects, while big changes unable to prime a specific threshold can see their effects dissipated. These phenomenal traits have been object of much interest in recent discussions on emergent properties. Processes of *amplification*, like the ones commonly exploited by cloud chambers or Geiger counters, are often mentioned as a key to conceptualise emergence (Deacon 2006; Bitbol 2007). In general, *non-linear processes* (Harth 2008: 61), with special reference to *positive or negative feedback* (Bickhard, Campbell 2000: 342) have been considered natural carriers of emergence. Here, however, we are *not* interested in arguing for any particular description of emergent processes, which could be open to controversy. What we want to show is that before or beyond any scientific evidence, the most ordinary phenomena, in a monistic framework, are enough to forcefully argue for *qualitative efficaciousness* and *discontinuities in the expression of such efficaciousness*; and these are traits that satisfy our requirements for emergent properties. Emergence here does not conflict with physical principles like the conservation of energy, because the novelty of emergent properties does not depend on the introduction of *further energy* in a closed system: the qualitative efficaciousness that we have outlined depends only on the fact that thresholds *select* and *modulate* already existing effects (Murphy 2006: 227). In classical terms, this could be expressed by recovering from the philosophical tradition the notion of *formal causation*. Such a formal causation must not be read as transferral of quantity, nor as succession of active causes and passive effects: here efficaciousness immediately depends on the *nature of co-occurring qualities*.

At this point we are entitled to say that room for emergence has been made. On the negative side, we have seen that the sciences of nature are not committed to any specific idea of causality and are therefore perfectly compatible with emergent properties. This clears the ground from the idea that there is a single legitimate model of natural efficaciousness against whose background all other models should seek justification. On the positive side, we have seen that sheer appearances in a monistic framework are sufficient to conclude that irreducible qualities must exist, that they must be efficacious and that a model of qualitative efficaciousness is available in terms of selective thresholds. At this point, the traditional burden of proof is reversed: we need not anymore provide special arguments to demonstrate, for instance, that logical or teleological motives are as efficacious as "ordinary" efficient causation is supposed to be. On the contrary, now it is up to the physicalist to demonstrate, if she can, that thoughts or motivations are either reducible resultants or ontologically inefficacious.

If we apply the above sketched model to the idea of the efficaciousness of the mind as mind (downward causation), we can tentatively draw its main features along these lines.

First, we need not believe that the mind creates anything ex-nihilo; in this sense we can grant that all that the mind has available is the "matter-energy" of the brain. Thus, mental processes *can* be said to be supervenient on "material" processes.

Yet, such supervenience does not legitimate any bottom-up model of causation. What is especially banned is the *causal continuity* between "lower" and "higher" levels, which makes plausible to infer from the supervenience of mind on cerebral matter to the unilateral production of mental events by changes in cerebral matter. Mind-brain supervenience shows just that matter and thought (or matter and life) belong to the same causal universe. When we acknowledge that material interventions on cerebral matter produce effects in the mental sphere, this should *not* be read as an "approximation", possibly to be refined by technological improvement, of the causal determination of the mind. It may well be that technical improvements in neurophysiologic interventions will enable us to thoroughly "cheat" a mind through cerebral changes, and by that means to induce more and more sophisticated *illusions elaborated by the mind*. But, regardless of how sophisticated such interventions may be, they could be never equivalent to "causing thoughts": any such "cheating" presupposes the living spontaneity of the "cheated" mind, which, insofar as it is a mind, *actively* deals with the raw matter of "bottom-up effects", while being irreducible to them. The essential point here is that not all physical causes operating on the body (or the brain) must or can *affect the mind as mind*. Some physical causes can *annihilate* the mind: they are not mental events. Some other physical causes can remain *unperceived* (in themselves and in their train of physical consequences): they are no mental events either. Further physical causes can produce "interferences" that cannot be *synthesised* by the mind, while being noticed: confusion, tiredness, etc. can be causally produced without obtaining any objective status among spatiotemporal phenomena. Finally, some physical causes become objects of consciousness positioned in the world: this is pre-eminently the case of "inner" and "outer" percepts (inclusive of illusions).

What we call "*activity* of the mind" is the selective and modulating *nature* of the mind. We must not introduce discussions about "mental causation" starting from what the mind *does* or can do, but from what the mind *is*. If we suspend the tacit obviousness that "the essence of causation is known" and if we make room for the intuition that *qualitative efficaciousness in the form of ontological thresholds* must exist, then the mind can supervene on cerebral matter without being either reduced to physical causes or doomed to impotence. The mind *shapes* efficaciousness, *selects* energy, *qualifies* causation in ways that may be said to be *potentially inherent in matter*, but are specifically actual only in the mental sphere.

Andrea Zhok
andrea.zhok@unimi.it
University of Milan

References

Alexander, S., 1920, *Space, Time, and Deity*, Macmillan, London.

Beckermann, A., Flohr, H., e Kim, J., eds., 1992, *Emergence or Reduction?*, Walter de Gruyter, Berlin.

Bedau, M., 1997, "Weak emergence", *Philosophical Perspectives*, Mind, Causation, and World, 11: 375-399

Bedau, M., 2008, "Is weak emergence just in the mind?", *Mind and Machines*, 18: 443-459.

Bickhard, M., Campbell, D.T., 2000, "Emergence", in Anderson P. *et al.*, eds., *Downward Causation: Minds, Bodies, and Matter*, Aarhus University Press, Aarhus: 322-348

Bitbol, M., 2007, "Ontology, matter and emergence", *Phenomenology and Cognitive Sciences*, 6: 293-307

Broad, C.D., 1925, *The Mind and Its Place in Nature*, Routledge & Kegan Paul, London.

Block, N., 1997, "Anti-Reductionism slaps back", *Philosophical Perspectives*, 11: 107-133.

Block, N., 2011, "Functional reduction", in Horgan, T. *et al.*, eds., *Supervenience in Mind. A Festschrift for Jaegwon Kim*, Mit Press, Cambridge MA.

Carnap, R., Morris, C., Neurath, O., eds., 1938-1969, *International Encyclopedia of Unified Science: Foundations of the Unity of Science*, University of Chicago Press, Chicago.

Clayton, Ph., Davies, P., eds., 2006, *The Re-Emergence of Emergence*, Oxford University Press, Oxford.

Clayton, Ph., 2006, "Conceptual foundations of emergence theory", in Clayton and Davies 2006: 1-33.

Craver, F.C., Bechtel, W., 2007, "Top-down causation without top-down causes", *Biology and Philosophy,* 22: 547-563.

Deacon, T., 2006, "Emergence: the hole at the wheel's hub", in Clayton and Davies 2006: 111-150.

Dowe, P., 2000, *Physical Causation*, Cambridge University Press, New York.

Fair, D., 1979, "Causation and the flow of energy", *Erkenntnis*, 63: 219-250.

Harth, E., 2008, "The element of time in the emergence of mental phenomena", *Journal of Consciousness Studies*, 15, 4: 54-65.

Hegel, G.W.F., 1999, *Wissenschaft der Logik*, Hauptwerke Bd. 3, Felix Meiner Verlag, Hamburg.

Husserl, E., 1913, *Logische Untersuchungen. Bd. I. Prolegomena zur reinen Logik*, Max Niemeyer Verlag, Halle.

Huxley, T.H., 1874, "On the hypothesis that animals are automata, and its history", *The Fortnightly Review*, 16 (New Series): 555-580.

Kant, I., 1998, *Critique of Pure Reason*, eng. tr. by P. Guyer, and A. Wood, Cambridge University Press.

Kim, J., 1992, "'Downward Causation' in emergentism and nonreductive physicalism" in Beckermann F., Kim, J., (eds.) *Emergence or Reduction?*, de Gruyter, Berlin: 119-138.

Kim, J., 1997a, "Does the problem of mental causation generalize?", *Proceedings of the Aristotelian Society*, 97: 281-297.

Kim, J., 1997b, "Explanation, prediction, and reduction in emergentism", *Intellectica,* 25: 45-57.

Kim, J., 1998, *Mind in a Physical World*, MIT Press, Cambridge MA.

Kim, J., 1999, "Making sense of emergence", *Philosophical Studies*, 95: 3-36.

Kim, J., 2005, *Physicalism, or Something Near Enough*, Princeton University Press, Princeton.

Kim, J., 2006a, "Being realistic about emergence", in Clayton and Davies 2006: 189 ff.

Lewes, G.H., 1875, *Problems of Life and Mind*, vol. 2, Kegan Paul, Trench, Turbner, and Co, London.

Lloyd-Morgan, C., 1923, *Emergent Evolution*, Williams and Norgate, London.

McLaughlin, B.P., 1992, "The rise and fall of British emergentism", in Beckermann 1992: 49-93.

Murphy, N., 2006, "Emergence and mental causation", in Clayton and Davies 2006: 227-243.

Nagel, E., 1961, *The Structure of Science*, Harcourt, Brace and World, New York.

O'Connor, T., 1994, "Emergent Properties", *American Philosophical Quarterly*, 31: 91-104.

Rothschild, L., 2006, "The role of emergence in biology", in Clayton and Davies 2006: 151-165.

Salmon, W., 1998, *Causality and Explanation,* Oxford University Press, New York.

Soto, A.M., Sonnenschein, C., 2006, "Emergentism by default: A view from the bench", *Synthese,* 151: 361-376.

Sperry, R.W., 1991, "In Defense of Mentalism and Emergent Interaction," *Journal of Mind and Behavior,* 12, 2: 221-245.

Van Gulick, R., 2001, "Reduction, emergence and other recent options on the mind/body problem: A philosophic overview", *Journal of Consciousness Studies*, 8, 9-10: 1-34

Von Wright, G., 1971, *Explanation and Understanding*, Routledge, London.